Revista Colombiana de Estadística
July 2018, Volume 41, Issue 2, pp. 235 to 249
DOI: http://dx.doi.org/10.15446/rce.v41n2.65654

Kernel Function in Local Linear Peters-Belson Regression

Función del núcleo en la regresión lineal local de Peters-Belson

Mohammad Bolbolian Ghalibaf[a]

Department of Statistics, Faculty of Mathematics and Computer Science, Hakim Sabzevari Univercity, Sabzevar, Iran

Abstract

Determining the extent of a disparity, if any, between groups of people, for example, race or gender, is of interest in many fields, including public health for medical treatment and prevention of disease or in discrimination cases concerning equal pay to estimate the pay disparities between minority and majority employees. The Peters-Belson (PB) regression is a form of statistical matching, akin in spirit to Bhattacharya's bandwidth matching which is proposed for this purpose. In this paper, we review the use of PB regression in legal cases from Bura, Gastwirth & Hikawa (2012). Parametric and nonparametric approaches to PB regression are described and we show that in nonparametric PB regression a suitable kernel function can improve results, i.e. by selecting the appropriate kernel function, we can reduce bias and variance of estimators, also increase the power of tests.

***Key words*:** Kernel function; Local linear regression; Welch's approximation.

Resumen

Determinar el alcance de una disparidad, si la hubiere, entre grupos de personas, por ejemplo, raza o género, es de interés en muchos campos, incluida la salud pública para el tratamiento médico y la prevención de enfermedades o en casos de discriminación en relación con la igualdad salarial para estimar las disparidades salariales entre los empleados minoritarios y mayoritarios. La regresión de Peters Belson (PB) es una forma de coincidencia estadística, similar en espíritu a la coincidencia de ancho de banda de Bhattacharya que se propone para este propósito. En este trabajo, repasamos el uso de la regresión del PB en casos legales de Bura et al. (2012). Se describen los enfoques paramétricos y no paramétricos de la regresión del

[a]PhD. E-mail: m.bolbolian@hsu.ac.ir; m.bolbolian@gmail.com

PB y demostramos que en la regresión no paramétrica del PB una función de kernel adecuada puede mejorar los resultados, es decir, seleccionando la función de kernel apropiada, podemos reducir el sesgo y la varianza de los estimadores, también aumentan el poder de las pruebas.

***Palabras clave*:** Aproximación de Welch; función kernel; regresión lineal local.

1. Introduction

In disparate treatment and equal pay discrimination cases, the salaries of minority/disadvantaged group (DG) members should be compared to those of similarly qualified majority/advantaged group (AG) employees. The PB regression, as well as ordinary regression with an indicator representing group status, have been accepted by courts (Gray 1993). As Bura et al. (2012) point out "the PB method offers some advantages compared to standard regression with a dummy or indicator variable. First, the method is intuitive and comparatively easy to understand for a general audience (e.g., judges, juries, etc.). For example, in the context of sex discrimination cases, PB regression estimates the salary equation for the male group (AG) incorporating related covariates and then takes the difference between the female group's (DG) actual salary and the estimated salary that the DG employee would have received if s/he were paid according to the equation for AG employees. Moreover, the estimated pay differential obtained from the PB approach is individualized for each member of the protected group. In contrast, ordinary least squares linear regression with an indicator variable estimates a common overall effect of being DG, after adjusting for the relevant covariates. This approach assumes that any differential is the same across the entire range of covariate values. Another advantage of the PB method is that the females whose salaries are higher than predicted and were not discriminated against are readily identifiable".

The PB regression method was first introduced by Peters (1941) and Belson (1956) for conducting treatment-control comparisons that accounted for relevant covariates by creating statistical matches for the treatment group observations. Blinder (1973) and Oaxaca (1973) used this idea to decompose the difference between the means of the two groups into components. Gastwirth & Greenhouse (1995) applied the PB method to salary data as well as to logistic regression for binary responses in order to analyze the data arising from a case involving promotion decisions (*Capaci v. Katz and Besthoff*[1]). Furthermore, Nayak & Gastwirth (1997) extended the method to generalized linear models. Hikawa, Bura & Gastwirth (2010*a*) introduced nonparametric PB in regressions with a binary response as an alternative to logistic regression. Hikawa, Bura & Gastwirth (2010*b*) introduced the local linear regression in PB. They considered the unknown functions for modeling the mean response in the two groups. Then they used Epanechnikov kernel to estimate unknown functions (for details about linear and nonlinear regression see Achcar & Lopes, 2016).

[1] *Capaci v. Katz and Besthoff* 711 F. 2d 647, 5th Cir., 1983

In this paper, we review the use of PB regression in legal cases from Bura et al. (2012). We suggest using another kernel functions and show that choose an appropriate kernel function can improve the results. The layout of the paper is as follows: In Section 2, we review parametric PB regression, based on parametric ordinary linear regression. Section 3, introduces a recent nonparametric version from Hikawa et al. (2010*b*) that increases the applicability of the PB approach. In Section 4, we apply all of the methods outlined to data from a sex discrimination case, and Section 5 contains the simulation study. We present the conclusion in Section 6.

2. Parametric Peters-Belson Regression

In this section, we review parametric PB regression. Like the study of Bura et al. (2012), we assume that the salaries (Y) are determined by a set of covariates (e.g., seniority, education, etc.) plus normally distributed random errors (ε). Suppose the salaries for minority and majority employees are given, respectively, by

$$\begin{aligned} Minority(DG): &\quad Y_{1i} = X_{1i}^T\beta_1 + \varepsilon_{1i}, \quad i = 1, \dots, n_1 \\ Majority(AG): &\quad Y_{2j} = X_{2j}^T\beta_2 + \varepsilon_{2j}, \quad j = 1, \dots, n_2 \end{aligned}$$

X denotes the covariate vector and β the corresponding coefficient vector. The errors, in each equation, are assumed to be normally distributed with mean zero and variance σ_1^2 and σ_2^2, respectively. If $\beta_1 = \beta_2$, there is no unjustifiable pay differential and the difference of salaries is due to random variability. A meaningful measure of pay differential against the minority employee with a given value of the covariate is:

$$\delta_i = X_{1i}^T\beta_1 - X_{1i}^T\beta_2$$

If δ_i is negative, the i-th minority employee is underpaid compared to a majority employee with the same covariate values. In parametric PB, a linear regression model is fitted to the data for the majority employees. Then each minority member's salary is predicted by $X_{1i}^T\hat{\beta}_2$, where X_{1i} is the covariate vector for the i-th minority member and $\hat{\beta}_2$ is the least squares estimate of β_2. The difference, $D_i = Y_{1i} - X_{1i}^T\hat{\beta}_2$, between the actual and predicted salaries is the estimate of the pay differential of the i-th minority employee relative to a similarly qualified majority employee. Thus, D_i is the parametric PB estimate of δ_i. When the model is correct, D_i is unbiased for δ_i and the corresponding unbiased estimator for the average disparity overall minority employees

$$\delta = \frac{1}{n_1}\sum_{i=1}^{n_1}[X_{1i}^T\beta_1 - X_{1i}^T\beta_2] = \bar{X}_1^T(\beta_1 - \beta_2)$$

is equal to

$$\bar{D} = \frac{1}{n_1}\sum_{i=1}^{n_1} D_i = \frac{1}{n_1}\sum_{i=1}^{n_1}[Y_{1i} - X_{1i}^T\hat{\beta}_2],$$

where $\bar{X}_1$ is the mean vector of the minority covariate values. The variance of $\bar{D}$ is

$$Var(\bar{D}) = \frac{\sigma_1^2}{n_1} + \sigma_2^2 \bar{X}_1^T (X_2^T X_2)^{-1} \bar{X}_1, \tag{1}$$

where $X_2 = (X_{21}, X_{22}, \ldots, X_{2n_2})^T$ is the usual design matrix of the majority group. When we assume $\sigma_1^2 = \sigma_2^2 = \sigma^2$, the test statistic

$$t = \frac{\bar{D}}{\sqrt{\hat{\sigma}^2 [1/n_1 + \bar{X}_1^T (X_2^T X_2)^{-1} \bar{X}_1]}} \tag{2}$$

can be used to test the null hypothesis $\delta = 0$. Gastwirth (1989) suggested that majority observations be used to estimate the common variance σ^2 because under the hypothesis of no discrimination both majority and minority are supposed to be paid under the same system and hence the variances are supposed to be the same as well. If we use $\hat{\sigma}^2 = \hat{\sigma}_2^2$, under the null hypothesis of $\delta = 0$, the test statistic in (2) is t-distributed with $n_2 - p_2$ degrees of freedom, where p_2 is the number of parameters (coefficients) in the majority model.

Gastwirth (1989) discusses the form of the variance of $\bar{D}$ in simple regression and the hypothesis testing for δ. Nayak & Gastwirth (1997) focus on a slightly different version of δ and its estimator and derive its distributional properties.

When the error variances are assumed to be different, we can approximate the distribution of the test statistic by using Welch's approximation approach (Welch (1949); Scheffe (1970); Nayak & Gastwirth (1997)). Under this approximation, the test statistic distribution under the null is approximated by a t distribution with degrees of freedom:

$$df = \frac{\left(\frac{\sigma_1^2}{n_1} + \bar{X}_1^T (X_2^T X_2)^{-1} \bar{X}_1 \sigma_2^2 \right)^2}{\sigma_1^4 / [n_1^2 (n_1 - p_1)] + \sigma_2^4 / (n_2 - p_2) \bar{X}_1^T (X_2^T X_2)^{-1} \bar{X}_1 \bar{X}_1^T (X_2^T X_2)^{-1} \bar{X}_1} \tag{3}$$

Since σ_1^2 and σ_2^2 are unknown, the degrees of freedom are estimated by substituting $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ in (3). For more details see Hikawa (2009).

***Note* 1.** The measure of average pay differential δ was used by Gastwirth (1989) with the particular intention to analyze the data arising from pay discrimination cases where the two underlying mean salary lines do not cross each other in the range of covariate values that are of interest. When the two mean lines cross each other, some values of δ_i become negative and others become positive; as a result, taking the average will cancel out these negative and positive values and d will no longer be a meaningful measure of pay differential. Therefore, δ should be used only for the cases where the two mean lines do not cross. While this is a theoretical possibility, but this situation does not happen in practice. Therefore, we recommend that the regression lines be plotted and if they cross, try to find out why; perhaps a variable has been omitted. After considering all the necessary variables, we probably can be run the PB method.

3. Local Linear Peters-Belson Regression

In this section, we review the local linear regression technique to estimate the expected minority responses from the majority data in the PB approach (Bura et al. 2012). Then, we express the role of the kernel function in fitting the local linear regression model.

Hikawa et al. (2010*b*) mentioned two problems an analyst of pay discrimination data often encounters. The first is the difficulty in estimating the salary equation when it does not appear to follow any usual parametric forms (e.g., linear, quadratic). The second problem pertains to determining who the relevant male/majority employees are to be compared against female/minority employees of interest. Including too many irrelevant majority employees in the comparisons (e.g., male employees who are too senior compared to the target female employees) may introduce serious bias in the estimated disparity (Greiner 2008).

To address these problems, Hikawa et al. (2010*b*) introduced the local linear regression in PB. Local linear regression fits a linear regression in the neighborhood of the covariate values of each minority member. The method is well suited for equal pay cases since the estimation/prediction of the salary of a minority employee is based on majority employees whose qualifications are closest to those of the minority employee and thus should receive the greatest weight. Furthermore, the similarity of this method to matched-pairs is expected to make the results more understandable to judges and juries. Local linear regression is similar in spirit to bandwidth matching introduced by Bhattacharya (1989). However, the weight given to each majority observation decreases with the distance of the covariate values from the target minority member.

Like the study of Bura et al. (2012), suppose we have d covariates and the data consist of n_1 minority observations, $(X_{11}, Y_{11}), \dots, (X_{1n_1}, Y_{1n_1})$, and n_2 majority observations, $(X_{21}, Y_{21}), \dots, (X_{2n_2}, Y_{2n_2})$, where X is a vector of d fixed covariate values. The response values of minority and majority members are generated by the following equations:

$$\begin{aligned} Minority: &\quad Y_{1i} = m_1(X_{1i}) + \varepsilon_{1i}, \quad i = 1, \dots, n_1 \\ Majority: &\quad Y_{2j} = m_2(X_{2j}) + \varepsilon_{2j}, \quad j = 1, \dots, n_2 \end{aligned}$$

where $m_1(X_{1i}) = E(Y_1|X = X_{1i})$ and $m_2(X_{2j}) = E(Y_2|X = X_{2j})$. Also ε_{1i}'s and ε_{2j}'s are iid $N(0, \sigma_1^2)$ and $N(0, \sigma_2^2)$, respectively. The only assumption we make on the unknown functions modeling the mean response in the two groups, $m_1(X)$ and $m_2(X)$, is that they are twice differentiable.

As in the parametric PB definition of disparity, the pay disparity for the i-th minority member is

$$\delta_i = m_1(X_{1i}) - m_2(X_{1i}),$$

and the average disparity of all minority members is

$$\delta = \frac{1}{n_1} \sum_{i=1}^{n_1} [m_1(X_{1i}) - m_2(X_{1i})].$$

Let

$$Z_i = \begin{pmatrix} 1 & (X_{211} - X_{11i}) & \dots & (X_{2d1} - X_{1di}) \\ 1 & (X_{212} - X_{11i}) & \dots & (X_{2d2} - X_{1di}) \\ \vdots & \vdots & & \vdots \\ 1 & (X_{21n_2} - X_{11i}) & \dots & (X_{2dn_2} - X_{1di}) \end{pmatrix}$$

and $W_i = diag(W(||X_{2j} - X_{1i}||/h))$, where W is a kernel weight function and $||.||$ is a norm. Denoting the elements of the first row of $(Z_i^T W_i Z_i)^{-1} Z_i^T W_i$ by $S_{i1}, S_{i2}, \dots, S_{in_2}$, the fitted value for the design point X_{1i} is given by

$$\hat{m}_2(X_{1i}) = \sum_{j=1}^{n_2} S_{ij} Y_{2j}.$$

The estimated pay differential for the i-th minority member is

$$D_i = Y_{1i} - \hat{m}_2(X_{1i}) = Y_{1i} - \sum_{j=1}^{n_2} S_{ij} Y_{2j}.$$

The estimated average pay differential against all minority members and its variance are

$$\begin{aligned} \bar{D}_{LOC} &= \frac{1}{n_1} \sum_{i=1}^{n_1} D_i = \frac{1}{n_1} \sum_{i=1}^{n_1} \left(Y_{1i} - \sum_{j=1}^{n_2} S_{ij} Y_{2j} \right), \\ var(\bar{D}_{LOC}) &= \frac{1}{n_1^2} \left(n_1 \sigma_1^2 + \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} S_{ij}^2 \sigma_2^2 + \sum_{i \neq l}^{n_1} \sum^{n_1} \sum_{j=1}^{n_2} S_{ij} S_{lj} \sigma_2^2 \right). \end{aligned}$$

Since σ_1^2 and σ_2^2 are usually unknown, the estimated variance of $\bar{D}_{LOC}$ can be obtained by using estimates $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ which can be obtained from the residuals of the separate local linear regression models within each group. Details of the estimation approach are given in Hikawa (2009) and Hikawa et al. (2010*b*). The estimated variance of $\bar{D}_{LOC}$ is given by

$$\widehat{var}(\bar{D}_{LOC}) = \frac{1}{n_1^2} \left(n_1 \hat{\sigma}_1^2 + \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} S_{ij}^2 \hat{\sigma}_2^2 + \sum_{i \neq l}^{n_1} \sum^{n_1} \sum_{j=1}^{n_2} S_{ij} S_{lj} \hat{\sigma}_2^2 \right).$$

Cleveland & Devlin (1988) showed that, under the assumption of normal errors and negligible bias of $\hat{m}_i(X)$, the distribution of $(\eta_j^2 \hat{\sigma}_j^2)/(\kappa_j \sigma_j^2)$ can be approximated by a χ^2 distribution with degrees of freedom η_j^2/κ_j, where

$$\begin{aligned} \eta_j &= tr(I - S_j)(I - S_j)^T \\ \kappa_j &= tr[(I - S_j)(I - S_j)^T]^2 \end{aligned}$$

for $j = 1, 2$. Let S_j be the $n_j \times n_j$ matrix whose $(i, k)^{th}$ element is S_{ik} obtained from fitting a separate smooth curve for the minority and majority groups (i.e., separately estimating $m_1(X_i)$ for $i = 1, \dots, n_1$ and $m_2(X_j)$ for $j = 1, \dots, n_2$). The test statistic for testing for lack of disparity ($H_0 : \delta = 0$ vs. $H_1 : \delta \neq 0$) is defined to be

$$t = \frac{\bar{D}_{LOC}}{\sqrt{\widehat{var}(\bar{D}_{LOC})}}. \tag{4}$$

When the two variances are equal (i.e., $\sigma_1^2 = \sigma_2^2 = \sigma^2$), we estimate the common variance by the majority variance estimate $\hat{\sigma}^2 = \hat{\sigma}_2^2$; see Gastwirth (1989). The estimated variance of $\bar{D}_{LOC}$ becomes

$$\widehat{var}(\bar{D}_{LOC}) = \frac{\hat{\sigma}^2}{n_1^2}\left(n_1 + \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} S_{ij}^2 + \sum_{i \neq l}^{n_1}\sum^{n_1}\sum_{j=1}^{n_2} S_{ij}S_{lj}\right).$$

Therefore, under the assumption of equal variances, the test statistic for group disparity is t distributed with $\frac{\eta^2}{\kappa}$ degrees of freedom, where $\eta = \eta_2$ and $\kappa = \kappa_2$.

When the two variances are assumed to be different, as in parametric PB, we can apply Welch's approximation approach to find the approximate distribution of the test statistic. The expression of the variance of $\bar{D}_{LOC}$ can be expressed as

$$\begin{aligned} var(\bar{D}_{LOC}) &= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1^2}\left(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} S_{ij}^2 + \sum_{i \neq l}^{n_1}\sum^{n_1}\sum_{j=1}^{n_2} S_{ij}S_{lj}\right) \\ &= \sigma_1^2 C_1 + \sigma_2^2 C_2. \end{aligned}$$

When H_0 is true, the test statistic (4) is approximately t distributed with the degrees of freedom

$$df = \frac{(\sigma_1^2 C_1 + \sigma_2^2 C_2)^2}{\sigma_1^4 \kappa_1 / \eta_1^2 C_1^2 + \sigma_2^4 \kappa_2 / \eta_2^2 C_2^2}. \tag{5}$$

Since σ_1^2 and σ_2^2 are unknown in most practical situations, the degrees of freedom are approximated by plugging $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ into (5).

The nearest neighbor bandwidth that fixes the fraction of data that contribute to the estimation (Cleveland & Devlin (1988); Loader (1999)) is used in fitting the local linear regression model. Hikawa et al. (2010*b*) is used Epanechnikov kernel as the weight function, but we recommend to use another kernels. The kernel function $W(u)$ is a non-negative real-valued integrable function satisfying the following requirements:

(i) $W(u) \geq 0$ for all u;

(ii) $\int_{-\infty}^{\infty} W(u)du = 1$;

(iii) $W(.)$ is symmetric about zero;

In Table 1 some samples of kernels are given. In these kernels, $u = ||X_{2j} - X_{1i}||/h$ and h is the bandwidth. Epanechnikov kernel is the optimal kernel in the sense that roughly both the asymptotic mean squared error and the mean integrated squared error are minimized (Fan & Gijbels 1996). Loader (1999) suggests scaling the vector norm.

Table 1: Some Kernel Functions.

Kernel	$\mathbf{W(u)}$
Epanechnikov	$\frac{3}{4}(1-u^2)I(\lvert u\rvert \leq 1)$
Normal	$\frac{1}{\sqrt{2\pi}}e^{\frac{-u^2}{2}}$
Logistic	$\frac{1}{e^u+e^{-u}+2}$
Laplace	$\frac{1}{2}e^{-\lvert u\rvert}$

We use

$$||X_{2j} - X_{1i}||^2 = \sum_{k=1}^{d} \left(\frac{X_{2kj} - X_{1ki}}{S_k} \right)^2,$$

where

$$S_k = \sqrt{\frac{\sum_{j=1}^{n_2}(X_{2kj} - X_{1ki} - \sum_{j=1}^{n_2}(X_{2kj} - X_{1ki})/n_2)^2}{n_2 - 1}}.$$

Since X_{1ki} is fixed, S_k is simply the sample standard deviation of X_{2k} about X_{1ki} for the majority members for the k-th covariate.

But using Epanechnikov kernel cannot be optimized, since in Epanechnikov kernel if $|u| \leq 1$ then $W(u) = 0$, while in practice, there may be some values of $|u|$ that exceed from 1 and thus this kernel doesn't have correctly results to estimate of these values. Therefore, we recommend using the kernel functions that values of u belong to real numbers. We used Normal kernel, Logistic kernel, and Laplace kernel and we concluded these kernels reduce bias and variance of estimators and increase the power of the test. In this paper, we will call these real kernels.

4. Application: Re-analysis of Data from *EEOC v. Shelby County Government*

In this section, we display the role of kernel functions in local linear PB regression by using data set from the pay discrimination case, *EEOC v. Shelby County Government*. This data set reported and analyzed by Bura et al. (2012). In 1983, a female clerical employee of Shelby County Criminal Court brought a charge of salary discrimination. The data, presented in Table 2, consist of salaries per pay period as the response and seniority measured in months since hired as the covariate with 16 male and 15 female clerical employees as of February 1, 1984. Fitting a ordinary regression to each gender separately yields:

$$\hat{Y}_f = 885.103 + 4.328X_f \tag{6}$$

$$\hat{Y}_m = 913.897 + 5.953X_m \tag{7}$$

where the subscripts m and f stand for male and female. The estimated variances of the error terms are $\hat{\sigma}_m^2 = 7494.5$ and $\hat{\sigma}_f^2 = 8757.1$, and the sample means and variances of seniority are: $\bar{X}_m = 81$, $S_m^2 = 1603.3$, $\bar{X}_f = 69.3$, and $S_f^2 = 2195.5$. Table 3 summarizes the ANOVA tables for both the male and female models.

TABLE 2: Shelby County pay discrimination case data.

Male		Female	
Salary (Y_2)	Seniority (X_2)	Salary (Y_1)	Seniority (X_1)
1666	125	1474	128
1666	124	1474	96
1666	124	1403	121
1666	120	1403	117
1548	117	1403	114
1548	116	1336	89
1548	113	1336	97
1548	95	1157	105
1548	73	1000	64
1306	69	1000	51
1157	51	1000	13
1157	47	1000	17
1157	46	929	13
1157	18	929	8
1000	41	929	6
1000	17		

TABLE 3: ANOVA tables for male and female models from the Shelby County data.

	Male				Female			
Sources	SS	DF	MS	F	SS	DF	MS	F
Seniority	852412	1	852412	113.74*	575666	1	575666	65.74*
Error	104924	14	7495		113842	13	8757	
Total	957336	15			689508	14		
	$R^2 = 0.89$				$R^2 = 0.83$			

*P-value was smaller than .0001.

Since the sample sizes of 16 males and 15 female are too small to fit a local linear regression model, we augmented the data set following the method used by Bhattacharya (1989) and Bhattacharya & Gastwirth (1999), where they analyze data from *Berger v. Iron Workers Local 201*. Fitting a Gamma distribution to the seniority data yielded Gamma(4, 20) for males and Gamma(2, 33) for females. Then, we consider two scenarios for enhancing observations. The first, we generate additional salary data for 34 males and 35 females according to the fitted models in (6) and (7) and other, we generate additional salary data for 84 males and 85 females according to the fitted models in (6) and (7). Therefore, in two scenarios sample sizes of male and female are same and equal to 50 and

100, respectively. The error variances were set equal to $\hat{\sigma}_m^2 = 7494.5$ and $\hat{\sigma}_f^2 = 8757.1$ in order to match the estimated variances of the error terms from the fitted regression models. The data were simulated based on unequal variances for males and females. Consequently, we compute the variance of $\bar{D}$ based on the assumption of unequal variances and approximate the degrees of freedom of the test statistics using the Welch's approximation approach. Tables 4 and 5 summarize the results from applying the five methods: (1) Parametric PB, (2) Local Linear PB with Epanechnikov kernel, (3) Local Linear PB with Normal kernel, (4) Local Linear PB with Logistic kernel and (5) Local Linear PB with Laplace kernel.

In Tables 4 and 5, the negative values of $\bar{D}$ indicate that female employees were underpaid on average compared to their similarly qualified male counterparts. The bias and standard error of the average pay differential estimated. In Table 4, value of $\bar{D}$ of the local linear PB with Epanechnikov kernel is different with other methods and its variance is greater than the other local linear methods. However, the parametric PB has the minimum variance. In Table 5, the results of methods are similar, but parametric PB still has the least variance and among nonparametric methods, the real kernels have fewer variances. These results demonstrate using the appropriate kernel can be effective in reducing bias and standard error. However, from a single example, one cannot make general conclusions. Therefore, we conducted a further simulation study, discussed in the next section.

Table 4: Analysis of the augmented Shelby County Pay Discrimination data ($n = 50$).

	Parametric PB	Local PB (Epanechnikov)	Local PB (Normal)	Local PB (Logistic)	Local PB (Laplace)
$\bar{D}$	-133.979	-129.116	-133.029	-133.206	-133.833
Bias	4.572	9.435	5.522	5.345	4.718
$\sqrt{\widehat{var}(\bar{D})}$	16.203	21.349	19.893	18.707	18.954
$\bar{D}/\sqrt{\widehat{var}(\bar{D})}$	-8.267	-6.048	-6.687	-7.121	-7.061
P-value	<.001	<.001	<.001	<.001	<.001

Table 5: Analysis of the augmented Shelby County Pay Discrimination data ($n = 100$).

	Parametric PB	Local PB (Epanechnikov)	Local PB (Normal)	Local PB (Logistic)	Local PB (Laplace)
$\bar{D}$	-125.542	-122.019	-122.364	-122.480	-123.732
Bias	1.500	5.023	4.678	4.562	3.310
$\sqrt{\widehat{var}(\bar{D})}$	11.625	15.625	14.373	13.791	14.017
$\bar{D}/\sqrt{\widehat{var}(\bar{D})}$	-10.799	-7.809	-8.513	-8.879	-8.827
P-value	<.001	<.001	<.001	<.001	<.001

Since the estimated amount of pay differential from all the methods is quite large, the p-values of all test statistics are very small. Hence, all methods would reject the null hypothesis of no pay differential and confirm the court's conclusion that the female employees were discriminated in their pay with an average differential about $129-134 ($122-125) per pay period for sample size 50 (100).

5. Simulation

In this section, we display the role of kernel functions in local linear PB regression by using simulated data. It should be noted that the simulation models were chosen from Bura et al. (2012). Consider a company that has three stores with employees of both sexes. In the first, amount of pay is same for two group (male and female). The data of this store simulated from equation (8).

$$y_i = 20000 + 200x_i + \varepsilon_i. \tag{8}$$

In the stores 2 and 3, amount of pay are different for two group and women are underpaid relative to comparable men in both stores but the system leading to the disparity differ in the two stores. In the second store, men and women start at the same salary but men receive better raises over time, while in the third store men start at a slightly higher salary and also receive higher raises over time. The data of these two stores simulated from equations (9) and (10), respectively.

$$y_i = \begin{cases} 20000 + 200x_i + \varepsilon_{1i} & \text{when } i \text{ is female} \\ 20000 + 250x_i + \varepsilon_{2i} & \text{when } i \text{ is male.} \end{cases} \tag{9}$$

$$y_i = \begin{cases} 20000 + 200x_i + \varepsilon_{1i} & \text{when } i \text{ is female} \\ 20500 + 250x_i + \varepsilon_{2i} & \text{when } i \text{ is male.} \end{cases} \tag{10}$$

The values of the seniority predictor variable for females and males were generated from the Gamma distribution with scale parameters 3 and 2 for females and males, respectively, and shape 2 for both sexes. To sum up, the simulations run according to 3 choices for sample sizes of two groups. The first, we choose in all three stores, the number of males is 40 and the number of females is 30, in the second choice, we consider the number of males is 80 and the number of females is 60 and the last scenario is based on 60 males and 80 females. The error variances were the same and errors generated from the normal distribution with mean 0 and standard error 300. Ten thousand replicates were used in the simulation and the significant level of $\alpha = 0.05$ considered.

Table 6 reports the results of the first choice for sample sizes. In the first store, where amount of pay is equal for two group, by using parametric PB the null hypothesis $\delta = 0$ is rejected 5.75% of the times, But by using local PB with Epanechnikov kernel the null hypothesis is rejected 5.96% of the times and by real kernels Normal, Logistic and Laplace the null hypothesis is rejected 5.35%, 5.28% and 5.34% of the times, respectively. Therefore, the size of all tests is close to nominal level 5%. In store 2, by using parametric PB the null hypothesis $\delta = 0$ is rejected 95.22% of the times, But by using local PB with Epanechnikov kernel the null hypothesis is rejected 81.97% of the times and by real kernels Normal, Logistic and Laplace the null hypothesis is rejected 86.40%, 93.47% and 93.09% of the times, respectively. The comparison of the power of the local linear tests shows that tests based on Logistic and Laplace kernels have the high powers followed by the Normal kernel, and Epanechnikov kernel, respectively. In store 3, by using parametric PB the null hypothesis $\delta = 0$ is rejected 100% of the times, But by using local PB with Epanechnikov kernel the null hypothesis is rejected 98.46% of the times and by real kernels Normal, Logistic and Laplace the null hypothesis is rejected 99.25%, 100% and 100% of the times, respectively. The comparison of the power of tests show that all tests have good powers. Furthermore, in all three stores, the standard error of local PB with Epanechnikov kernel is more than other methods.

Table 6: Analysis of the simulated data (Scenario 1).

		Parametric PB	Local PB (Epanechnikov)	Local PB (Normal)	Local PB (Logistic)	Local PB (Laplace)
Store 1	$\bar{D}$	-0.286	-1.235	1.342	-0.038	-0.038
	Bias	-0.286	-1.235	1.342	-0.038	-0.038
	Std. Dev.	81.129	109.472	101.126	84.725	85.438
	Size	0.0575	0.0596	0.0535	0.0528	0.0534
Store 2	$\bar{D}$	-300.692	-301.495	-298.997	-300.368	-300.386
	Bias	-0.539	-1.342	1.155	-0.215	-0.233
	Std. Dev.	81.054	107.611	100.987	84.592	85.291
	Power	0.9522	0.8197	0.8640	0.9347	0.9309
Store 2	$\bar{D}$	-799.564	-801.386	-798.246	-799.610	-799.584
	Bias	0.460	-1.362	1.778	0.415	0.441
	Std. Dev.	81.090	106.392	100.198	84.627	85.328
	Power	1	0.9846	0.9925	1	1

Tables 7 and 8 show the results of two other choices for sample sizes. The findings of these two tables are similar to the previous table. The size of all tests is close to nominal level 5%, and tests based on real kernels have the good powers. Also, using local PB with Epanechnikov kernel causes the bias and larger standard error.

Table 7: Analysis of the simulated data (Scenario 2).

		Parametric PB	Local PB (Epanechnikov)	Local PB (Normal)	Local PB (Logistic)	Local PB (Laplace)
Store 1	$\bar{D}$	0.254	-4.472	-0.033	0.147	0.123
	Bias	0.254	4.472	-0.033	0.147	0.123
	Std. Dev.	56.892	78.072	63.938	58.313	58.614
	Size	0.0546	0.0503	0.0521	0.0513	0.0514
Store 2	$\bar{D}$	-299.598	-302.356	-299.417	-299.658	-299.620
	Bias	0.243	-2.514	0.425	0.183	0.221
	Std. Dev.	56.850	76.537	64.224	58.318	58.628
	Power	0.9995	0.5765	0.9833	0.9988	0.9986
Store 2	$\bar{D}$	-800.173	-797.535	-799.796	-800.214	-800.214
	Bias	0.021	2.659	0.398	-0.019	-0.019
	Std. Dev.	56.852	78.295	64.217	58.285	58.591
	Power	1	0.8831	0.9997	1	1

Table 8: Analysis of the simulated data (Scenario 3).

		Parametric PB	Local PB (Epanechnikov)	Local PB (Normal)	Local PB (Logistic)	Local PB (Laplace)
Store 1	$\bar{D}$	0.613	3.658	1.571	0.567	0.605
	Bias	0.613	3.658	1.571	0.567	0.605
	Std. Dev.	58.661	85.658	68.744	60.673	61.062
	Size	0.0551	0.0434	0.0522	0.0528	0.0525
Store 2	$\bar{D}$	-301.931	-304.083	-300.321	-301.715	-301.647
	Bias	-1.610	-3.761	0.0003	-1.393	-1.325
	Std. Dev.	58.732	83.029	68.777	60.725	61.122
	Power	0.9985	0.5303	0.9671	0.9965	0.9957
Store 2	$\bar{D}$	-800.018	-802.577	-799.363	-800.007	-800.026
	Bias	0.167	-2.390	0.822	0.178	0.159
	Std. Dev.	58.766	84.116	69.087	60.809	61.201
	Power	1	0.8433	0.9993	1	1

6. Conclusion

In this paper, we show that by selecting an appropriate kernel, we can use local linear PB regression and the results of this method are similar to the parametric method. Therefore in most cases, due to the constraints of parametric PB regression, we can be used local linear PB regression. According to the results of data from *EEOC v. Shelby County Government* and simulation study we concluded that the use of local linear PB regression be effective in determining the extent of a disparity between groups and in local PB method using real kernels can reduce the bias and variance of the estimators and also increase the power of the test. By comparing the three real kernel functions we deduced Logistic and Laplace kernels have better results. By using these kernels the bias is the lowest and variance of the estimator is relatively small. Moreover, the power of the test is acceptable. But, the Normal kernel compared to two other real kernels has weaker results. However, all three real kernel in comparison with the Epanechnikov kernel have better performances. It should be noted that we investigate the performance of some other kernels like Triangular and Cosine and have concluded that results of these kernels are similar to the Epanechnikov kernel. Therefore, we recommend using real kernels in local linear PB regression.

Acknowledgement

The author would like to thank the reviewer for their valuable suggestions and comments.

[Received: January 2017 — Accepted: January 2018]

References

Achcar, J. A. & Lopes, S. R. C. (2016), 'Linear and Non-Linear Regression Models Assuming a Stable Distribution', *Revista Colombiana de Estadística* **39**(1), 109–128.

Belson, W. A. (1956), 'A technique for studying the effects of a television broadcast', *Applied Statistics* **5**, 195–202.

Bhattacharya, P. K. (1989), Estimation of treatment main effect and treatment-covariate interaction in observational studies using bandwidth-matching, Technical Report 188, Division of Statistics, University of California, Davis.

Bhattacharya, P. K. & Gastwirth, J. L. (1999), 'Estimation of the Odds-Ratio in an Observational Study Using Bandwidth-Matching', *Journal of Nonparametric Statistics* **11**, 1–12.

Blinder, A. S. (1973), 'Wage discrimination: reduced form and structural estimates', *Journal of Human Resources* **8**, 436–455.

Bura, E., Gastwirth, J. L. & Hikawa, H. (2012), The use of peters–belson regression in legal cases, *in* 'Nonparametric Statistical Methods And Related Topics: A Festschrift in Honor of Professor PK Bhattacharya on the Occasion of His 80th Birthday', World Scientific, pp. 213–231.

Cleveland, W. S. & Devlin, S. J. (1988), 'Locally-weighted regression: an approach to regression analysis by local fitting', *Journal of the American Statistical Association* **83**, 597–610.

Fan, J. & Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.

Gastwirth, J. L. (1989), 'A clarification of some statistical issues in Watson vs. Fort Worth Bank and Trust', *Jurimetrics Journal* **29**, 267–284.

Gastwirth, J. L. & Greenhouse, S. W. (1995), 'Biostatistical Concepts and Methods in the Legal Setting', *Statistics in Medicine* **14**, 1641–1653.

Gray, M. W. (1993), 'Can statistics tell us what we do not want to hear? The case of complex salary structures', *Statistical Science* **8**(2), 144–158.

Greiner, D. J. (2008), 'Causal inference in civil rights litigation', *Harvard Law Review* **122**, 533–598.

Hikawa, H. (2009), Local linear peters-belson regression and its applications to employment discrimination cases, Doctoral dissertation, Department of Statistics , George Washington University.

Hikawa, H., Bura, E. & Gastwirth, J. L. (2010*a*), 'Local Linear Logistic Peters-Belson Regression and its application in employment discrimination cases', *Statistics and its Interface* **3**, 125–144.

Hikawa, H., Bura, E. & Gastwirth, J. L. (2010*b*), Robust peters-belson type estimators of measures of disparity and their applications in employment discrimination cases, Technical report, Department of Statistics, George Washington University.

Loader, C. (1999), *Local Regression and Likelihood*, Springer-Verlag, New York.

Nayak, T. K. & Gastwirth, J. L. (1997), 'The Peters-Belson approach to measures of economic and legal discrimination', *Advances in the Theory and Practice of Statistics* pp. 587–601.

Oaxaca, R. (1973), 'Male-Female differentials in urban labor markets', *International Economic Review* **14**, 693–709.

Peters, C. C. (1941), 'A method of matching groups for experiments with no loss of populations', *Journal of Educational Research* **34**(8), 606–612.

Scheffe, H. (1970), 'Practical solutions of the Behrens-Fisher problem', *Journal of the American Statistical Association* **65**(332), 1501–1508.

Welch, B. L. (1949), 'Further note on Mrs. Aspin's tables and on certain approximations to the tabled function', *Biometrika* **36**(3/4), 293–296.